

DB, 가공 통한 활용 중요

이용 목적에 따른 데이터마이닝 기술이 핵심



박성수 / 삼성카드 전략정보팀 과장

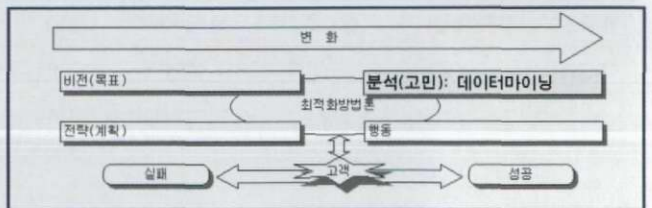
연재 순서

- 1 급변하는 시대와 CRM의 시대적 요청
- 2 CRM을 위한 기본적인 사상
- 3 고객과의 관계기반 '데이터웨어하우스'
- 4 부가가치정보를 위한 '데이터 마이닝'
- 5 마케팅의 효율화를 위한 '캠페인 매니지먼트 시스템'
- 6 전략정보 시스템으로의 'Risk Management'
- 7 컴퓨터와 전화의 통합 'CTI 콜센터'
- 8 고객을 만나는 또 하나의 채널 '인터넷'
- 9 콜센터와 인터넷의 만남 '인터넷 콜센터'
- 10 CRM의 성공을 위한 '코어시스템의 혁신'
- 11 성공적 CRM 구축을 위한 IT 방법론
- 12 CRM, 그 성공의 비결

앞서 제시했듯이 데이터웨어하우스(Dataware House)가 자료 창고라고 하면 데이터마이닝(Data mining)은 그 창고에 있는 데이터를 가지고 실제 분석을 하는 것이다. 창고에 있는 데이터 자체는 기업에 부가가치를 줄 수는 없다. 이것을 분석하여 유용한 자료가 나와야지 그것을 가지고 행동에 옮겨 기업은 이익을 얻기 때문이다. 예를 들면 아무리 야구에 대한 지식이 머리 속에 많은 감독이라도 고민을 하여 실제 경기에서 응용할 수 있는 작전이 없다면 그 지식은 의미가 없듯이 데이터가 아무리 훌륭하고 많아도 그것을 분석하는 단계에서 문제가 있다면 아무 소용이 없다는 것이다.

결론을 말하자면 머리 속에 있는 많은 지식을 데이터웨어하우스라고 말할 수 있고 이것을 가지고 뭔가 고민을 하는 행위를 데이터마이닝이라고 하는 것이다.

그렇다면 데이터마이닝만 하면 유용한 정보가 무조건 나오는 것일까?



그건 아니다. 이것은 다이아몬드를 만드는 것과 같은 이치다. 다이아몬드 가공업체를 보면 창고에 엄청 많은 원석(데이터웨어하우스)을 가지고 하나의 훌륭한 다이아몬드(유용한 자료)를 위해서 열심히 가공(데이터마이닝)을 한다. 그렇지만 열심히 가공을 한다고 전부 훌륭한 다이아몬드가 나오는 것은 아니다. 보기에는 변변하지 않은 원석(데이터)이지만 가공을 하고 나니까 훌륭한 다이아몬드가 될 수도 있고 겉은 훌륭한 원석(데이터)이지만 가공을 하니 별 볼일 없는 다이아몬드도 될 수 있다는 것이다.

정말 열심히 가공(데이터마이닝)을 했는데 별 볼일 없는 다이아몬드가 나오면 정말 속상하겠지만 그렇다고 훌륭한 다이아몬드(유용한 자료)가 나올 확률이 적어 가공할 필요가 없다고 생각하는 것은 올바른 생각일까?

유용하지 않는 자료가 나왔다고 데이터마이닝을 헛수고했다고 하면 그건 잘못된 생각이다. 기업이 데이터마이닝을 하지 않는다면 창고에 있는 데이터 중에 유용한 데이터가 뭔지, 별 볼일 없는 데이터가 뭔지 알 수 없기 때문이다. 만약 고객에게 어떤 행동으로 접근 할 때 별 볼일 없는 데이터를 기준으로 한다면 그건 오히려 큰 부작용을 일으킬 수도 있다. 또 데이터웨어하우스에 별 볼일 없는 데이터는 빨리 정리하여 유용한 정보로만 효율적으로 관리를 해야 하니 데이터마이닝의 행위 자체는 그것이 유용하든 안 하든 반드시 해야만 하는 것이다.

데이터마이닝의 정의

데이터마이닝을 한마디로 말한다면, 데이터베이스 내에 숨어 있는 일정한 패턴이나 변수들간의 관계를 정교한 분석모형을 사용하여 쉽게 드러나지 않는 유용한 정보를 찾아내는 과정이라고 할 수 있다.

예를 들면, '창문이 많은 집에는 도둑이 많다' 라는 말은 당연한 말처럼 들릴 수도 있지만, 누구나 이러한 사실이 보험회사에서 보험료를 계산하는데 적용될 수 있으리라고 생각하지는 않았을 것이다.

실제로 영국의 한 보험회사는 이러한 엉뚱한 사실을 이용하여 차별적인 보험 요율을 적용함으로써, 보다 효과적인 정책을 수립하게 되었고 이로 인해 기업의 수익증대에 커다란 공헌을 하게 되었다.

데이터마이닝은 다음과 같이 크게 Computer Science 관점

과 MIS 관점, Statistics 관점에 의한 정의로 나누어 살펴볼 수 있다. 먼저 Computer Science 관점은 패턴 인식 기술뿐 아니라 통계적, 수학적 분석방법을 이용하여, 저장된 거대한 자료로부터 우리에게 유익하고 흥미 있는 새로운 관계, 성향, 패턴 등 다양한 가치 있는 정보를 찾아내는 일련의 과정으로 정의하고 있다.

MIS 관점은 거대한 데이터 베이스 혹은 자료에서 유용한 정보를 유출하는 일련의 과정뿐 아니라 값진 정보를 사용자가 전문적 지식 없이 사용할 수 있는 의사 결정지원 시스템의 개발과정을 통틀어 데이터마이닝이라고 정의하고 있다.

Statistics 관점은 올바른 의사 결정을 지원하기 위한 자료분석(Data Analysis) 및 모델 선택 (Model Selection)으로 정의하고 있다. 기업이 보유하고 있는 일일 거래 데이터, 고객 데이터, 상품 데이터 혹은 각종 마케팅 활동에 있어서의 고객 반응 데이터 등과 이외의 기타 외부 데이터를 포함하는 모든 사용 가능한 근원 데이터를 기반으로 감춰진 지식, 기대하지 못했던 경향 또는 새로운 규칙 등을 발견하고 이를 실제 비즈니스 의사결정 등에 유용한 정보로 활용하고자 하는 것이 바로 데이터마이닝이다.

즉, 데이터마이닝은 다양한 방법을 이용하여 근원 데이터를 탐색하고 분석하여 이로부터 기대하고 있는 정보뿐만 아니라, 예상하지 못했던 정보를 찾아내고자 하는 개념적인 방법론인 것이다.

데이터마이닝의 특징

데이터마이닝의 특징은 크게 여섯 가지로 요약된다.

첫째, 데이터베이스에 비계획적으로 수집된 대용량의 데이터를 다룬다.

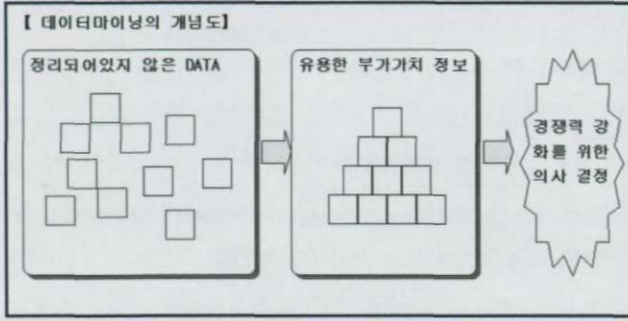
둘째, 컴퓨터의 강력한 처리능력을 이용한다.

셋째, 대다수의 데이터마이닝 기법들은 수학적으로 증명되고 발전된 것이 아니라 경험적으로 개발되었다.

넷째, 데이터마이닝의 주요 관심은 통계적 추론과 검정보다는 예측모형의 일반화에 있다.

다섯째, 기업의 다양한 의사결정 활동에 활용하기 위해서 사용된다.

여섯째, 데이터마이닝은 통계학, 전산과학, 인공지능, 공학 분야에서 개발되기 시작하였으나 실제로 경영, 경제, 정보기술



분야에서 활용되고 있다.

데이터마이닝을 이야기할 때, 'mining'이라는 것에만 초점을 두어 마치 특정기법(예를 들면, Neural Networks, Case Based Reasoning, Decision Tree 등)이 데이터마이닝이라고 잘못 이해하는 경우가 종종 있다. 그러나 데이터마이닝은 신경망모형(Neural Networks)이나 의사결정수(Decision Tree)와 같은 특정 기법을 말하는 것이 아니라 개념적인 정보추출의 방법론이며 이와 관련한 일련의 과정(Process)이라고 할 수 있다. 실제 데이터마이닝이 적용되는 프로세스를 살펴보면 다음과 같은 단계로 나누어 볼 수 있다.

첫째, 문제 정의 단계이다.

이는 적용하고자 하는 비즈니스 문제 정의 및 목표를 결정한 다. 즉 데이터마이닝의 필요성을 충분히 인식하고 현 비즈니스 문제에 대해 잘 이해하고 목적이 무엇인지를 확고히 할 필요가 있다. 정확한 문제의 이해 없이는 성공적인 데이터마이닝을 수행할 수가 없다. 또한 데이터마이닝의 결과로 얻어진 정보를 어떻게 활용할 것인가 하는 실제 업무와의 연계성도 충분히 고려해야 한다.

둘째, 데이터베이스 구축 단계이다.

이는 정의된 비즈니스 문제에 따라 필요한 데이터를 선택하고 데이터웨어하우스(또는 데이터마트)를 구축함으로써 데이터를 준비한다.

셋째, 데이터마이닝 과정단계이다.

즉 준비된 데이터를 샘플링하고, 사전분석을 통해 탐색하고 변형과정을 거친 후 적절한 데이터마이닝 기법을 이용하여 정보의 패턴을 발견하고 평가하는 것이다. 즉 보유하고 있는 데이터를 데이터마이닝을 할 수 있는 적정 상태로 준비하는 작업이다. 적용대상업무에 따라서 차이는 있겠지만 실제로 데이터마이닝 수행 시, 데이터가 제대로 준비되지 않은 상태에서 필요한 테이블을 연결하고 양질의 데이터를 준비하는 작업은 전체 과정의 80% 이상의 시간과 노력이 소비되기도 한다. 일단

작업대상이 되는 데이터가 준비되면 단순한 SQL을 이용한 질의나 OLAP, 여러 가지 그래픽적인 방법들(Visualization) 혹은 통계적인 방법들을 사용한 일종의 사전분석을 통해 데이터에 대한 기본적인 정보를 얻고 데이터를 이해하고 윤곽을 잡을 수가 있다. 이러한 탐색과정에서의 정보를 기반으로 하여 통계적인 방법들을 비롯한 적절한 데이터마이닝의 방법들이 적용된다.

넷째, 비즈니스 보고서작성 단계이다.

데이터마이닝 과정에서 얻어진 결과물에 대해 사용자가 쉽게 이해할 수 있도록 비즈니스의 문제와 목적에 맞게 재표현하는 단계라고 할 수 있다. 이러한 과정을 통해 얻어진 고급정보는 그 의미와 정도에 대해 해석하고 평가하는 단계를 거쳐 실제 업무의 목적에 적합한가를 판단한다. 실제 정보 사용자 혹은 의사결정자가 쉽게 이해할 수 있는 형태로 가공되어져 실제 업무에 활용될 수 있도록 결과물의 전달을 포함하는 사용자 환경을 구축하게 될 것이다.

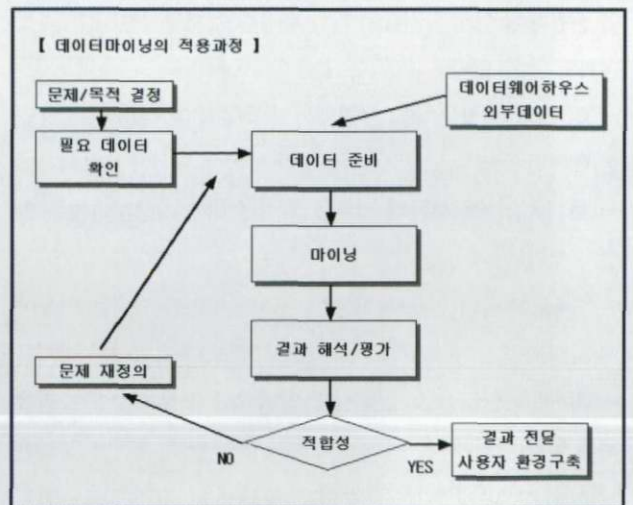
다섯째, 의사결정단계이다.

데이터마이닝으로부터의 정보를 기반으로 수립된 전략이나 의사결정을 통한 실제 업무에 있어 활용하는 단계이다.

여섯째, 피드백(Feedback)단계이다.

실제업무에서의 적용 후의 결과나 효과를 토대로 향상된 정보를 얻기 위해 데이터마이닝의 초기단계로 회귀하는 단계이다. 데이터마이닝을 통해 얻어진 정보는 실제 상황에서의 평가를 통해 피드백(Feedback)되어 다시 데이터마이닝에 반영되고 재분석이 되면서 얻어질 결과물의 신뢰성과 정도를 높여 가게 된다.

데이터마이닝은 의미와 목적상 의사결정시스템(Business



Intelligence System)과 병행되어야 한다. 즉, 데이터마이닝 솔루션은 특정업무에 국한되거나 단순히 데이터마이닝이 적용되기 위한 기법이나 방법론의 모음이 아니다.

다양하게 존재하는 원천데이터에 대한 용이한 접근이 가능해야 하고 유용하게 적용될 수 있는 여러 기법들을 제공할 수 있어야 한다. 어떤 문제를 다루는데 정해진 기법이나 규칙이 정해져 있는 것이 아니라 데이터에 따라 혹은 다루어야 할 문제의 성격에 따라 다양한 기법들이 적용될 수가 있어야 하기 때문이다.

또한 구체적인 mining의 방법론을 가지고 실제 작업을 구현하는데 있어서 지침이 되는 가이드라인이 제시되어야 하고 데이터마이닝에 의한 결과물을 최종사용자가 쉽게 이해하고 이용할 수 있도록 하는 부분까지를 포함하는 보다 포괄적인 솔루션이어야 한다.

그리고, 최근 기술의 발전과 더불어 여러 데이터마이닝 기법들이 소개되고 있지만, 데이터마이닝에서 인적요소의 역할은 매우 중요하다. 즉 사람에 의한 판단은 mining 과정에서 매우 중요하며, mining된 결과를 비교·평가하고 이를 실제 업무에 어떻게 활용할 것인가를 판단하는 것 역시 사람만이 가능하기 때문이다.

데이터마이닝의 주요 활용분야

데이터마이닝은 산업 전반에 걸쳐 많은 기업들은 그들이 구축한 세세한 거래정보 데이터 베이스에 데이터마이닝 기법을 적용하여 얻은 유용한 정보를 활용하므로써 경쟁우위를 확보하고 있다.

데이터마이닝이 활용되고 있는 분야들을 소개하면 다음과 같다.

마케팅(Marketing)

데이터마이닝이 가장 많이 활용되는 분야가 마케팅이다. 특히 최근에는 기업이 가지고 있는 고객정보의 데이터베이스를 중심으로 데이터마이닝 기법을 사용하여 데이터베이스 예측 모델의 개발하고 있다. 유통업자들은 고객의 구매행위, 예를 들어 어떤 상품의 구매행위나 할인 행사에 참여하는 행위 등을 통하여 특성을 파악할 수 있으며 이러한 지식을 통하여 특정 고객집단을 겨냥한 효과적이고 경제적인 판매 촉진 전략을 구사할 수 있다.

은행업 (Banking)

은행은 다음과 같이 사기행위 색출 (fraud detection), 고객 집단 분류(customer segmentation), 라이프사이클에 따른 고객 가치 관리 (predictive life cycle management) 등 다양한 분야에 데이터 마이닝을 이용할 수 있다.

1) 사기행위 색출 : 실제로 은행이 신용카드 사업에서 신용카드 사기로 인한 피해는 매우 크다. 이러한 신용카드의 사기적 발에 있어서 은행은 DataMining 기법을 이용하여 과거에 사기행위로 판명된 신용카드 거래를 분석하여 사기행위의 패턴을 찾아낼 수 있다. 신용카드 사기 행위의 전형적인 사례는 전자상가에서 짧은 기간에 많은 거래가 일어나는 경우이며 이것을 사기행위의 가능성을 알려 주는 경고 신호로 인식함으로써 피해를 줄일 수 있다. 또한 어떤 고객의 구매행위가 찾아낸 사기행위 패턴과 비슷할 경우 그 거래를 승인하지 않도록 시스템 (production system)을 구성하는데 이용할 수 있다.

2) 고객집단 분류 : 특정 고객집단을 찾아내고 이 집단만을 겨냥한 차별화된 서비스를 제공할 수 있다. 예를 들어, 어떤 금융 상품은 자주 해외여행을 다니는 고객집단에게, 어떤 상품은 언제나 결제일을 잘 지키는 고객들에게 중점적으로 판매할 수 있다. 또한 고객집단 편성에 관한 지식을 이용하여 특정 판촉활동에 의하여 가장 많은 효과와 혜택을 얻게 될 금융기관의 지점을 찾는 데에도 사용할 수 있다.

3) 라이프 사이클 예측 관리 (predictive life-cycle management) : 데이터마이닝은 은행이 고객의 시간에 따른 가치 (lifetime value)를 예측하고 이에 따라 개개의 고객집단에 알맞은 서비스를 제공하는데 도움을 준다. 은행은 현재의 수익성이 높은 고객집단을 정의하고 지식발견을 이용하여 이들의 몇 년 전의 공통된 특성을 발굴한다. 그 다음 이러한 특성을 지닌 현재의 고객들을 찾아낼 수 있는데 이들은 가까운 장래에 수익성이 높은 고객이 될 가능성이 매우 높은 고객들이다. 은행은 이들에게 특별한 상품 거래를 제안하거나 수수료를 면제해 주는 것과 같은 고객 이탈방지 프로그램 같은 것을 실시할 수 있다.

통신산업 (Telecommunications)

전 세계적으로 점점 치열해져 가는 경쟁에 직면하고 있는 통신회사들은 기존 고객을 유지하고 새로운 고객을 끌어들이기

위해 적극적인 마케팅 정책과 가격 정책을 실시하고 있다. 이러한 통신산업 분야에 지식발견이 적용된 예는 다음과 같다.

1)통화 기록 분석 : 통신사업자들은 고객의 자세한 통화 기록을 가지고 있다. 비슷한 통화 사용패턴을 가진 집단을 찾아내어 그들에게 유리한 가격정책이나 기능 등을 개발할 수 있다.

2)고객 충성도 (customer loyalty) : 어떤 고객은 계속 통신서비스 제공자를 바꾸면서 각 통신회사가 제공하는 인센티브를 이용한다. 통신회사는 지식발견 기술을 이용하여 한 번 고객이 되면 오랜 동안 지속적인 거래를 하게 될 고객과 그들의 특성을 찾아내고 이들을 중심으로 가장 이익이 많은 곳에 투자를 할 수 있다.

보험업 (Insurance)

보험회사는 보험이라는 상품의 특성상 오랜 기간에 걸쳐 집적된 방대한 데이터를 가지고 있는데 이것은 효과적인 계획을 세우는데 지렛대로 활용될 수 있다. 이러한 보험산업 분야에 지식발견 기술이 적용된 예는 다음과 같다.

1)사기 색출 : 보험회사는 예를 들어 교통사고 후유증과 같은 높은 보험 청구율을 가진 분야의 청구자, 의사, 변호사들 사이의 관련성 또는 보험청구 패턴을 찾아냄으로써 보험사기를 줄일 수 있다.

2)상품 설계 (product design) : 보험업자는 가장 수익성이 좋은 상품 구성 즉, 보험가입 신청자의 특성, 보험증권의 보장 범위, 및 보험증권의 특약의 최적의 결합을 알고 싶어한다. 보험업자들은 이 정보를 새로운 상품을 설계하고 장래의 판매를 위하여 기존의 상품을 고부가가치화 하는데 이용한다.

3)위험 분석 (risk analysis) : 보험업자는 보험지급액과 관련된 여러 요인들을 찾아냄으로써 지급부담 위험을 줄일 수 있다. 예를 들어 미국의 대형 보험회사는 최근 지난 2년 간의 중요한 보험 청구건을 검토한 결과 기혼자의 청구금액이 미혼자의 청구금액의 두 배에 달한다는 사실을 발견하였다. 이 지식을 바탕으로 이 회사는 기혼자에게 일률적으로 적용, 할인하여 주는 정책을 조정하였다.

그밖의 데이터마이닝 응용 사례

1)고객집단 편성 : 거의 모든 산업분야는 데이터마이닝을 이용하여 뚜렷한 고객집단을 편성하는데 활용함으로써 많은 이익을 얻을 수 있다. 이들 조직은 데이터마이닝을 이용하여 전통적

인 데이터웨어하우스에서 다루는 것보다 훨씬 많은 변수들을 고려할 수 있다.

2)자동차 제조 : 자동차 제조업자들은 고객들을 위하여 맞춤형 자동차를 생산하기 시작했고 따라서 어떤 특징들이 선호될 것인지 그리고 이러한 특징과 함께 어떤 점들이 요구될지를 예측해야 할 필요가 있다.

3)보증계약 (warranties) : 제조업자는 보상 청구를 해 올 고객의 수를 예측하고 이에 따른 비용을 예측하여야 한다.

4)탑승객 인센티브 (frequent flier incentives) : 항공사는 자사 비행기를 더 자주 이용할 수 있도록 인센티브를 제공할 고객 집단을 찾아내는데 데이터마이닝을 이용할 수 있다. 예를 들면 한 항공사는 비행거리 누적 혜택을 받는 데는 별 도움이 되지 않을 정도의 짧은 거리를 매우 자주 여행하는 고객집단이 존재함을 발견하였다. 따라서 이 항공사는 비행거리 뿐만 아니라 비행횟수에 의해서도 항공사가 제공하는 혜택을 받을 수 있도록 규칙을 변경하였다.

5)제조업(Manufacturing) : 공정 환경의 복잡함과 작업의 효율성과 품질의 우수성을 동시에 요구하는 현대 제조 분야에서도 데이터마이닝 기법이 여러 분야에서 필요시 되고 있다. Job shop scheduling, 공정 제어(manufacturing control)의 문제, 화학 약품의 공정 과정(chemical process)의 최적화, 에너지 소비(energy consumption)의 최소화 혹은 품질관리(quality control)와 자동화된 검사(automated inspection)에서도 데이터마이닝 기법이 쓰여지고 있다.

6)보건 의학 분야(Health and medical) : 이 분야에서의 데이터마이닝의 이용은 크게 두 가지로 나뉜다. 첫째는 환자나 병원 측에서의 보험 사기(fraudulent insurance claim)에 대한 검색과 가장 경제적인 의료제공원(cost-effective health care provider)을 판별해 내는데 쓰이고, 둘째는 이 기법으로 각종 암이나 심장마비 등의 병을 자동 진단(automated diagnosis)하는 데 있다.

7)에너지와 유틸리티(energy and utility) : 데이터마이닝을 이용하여 에너지의 수요량을 예측(forecast)하여 날씨의 변화와 정전에 신속하게 대비할 수 있으며, 또한 원유 탐사에 있어서 지층의 변화를 분석하는 데도 쓰여진다. ☞

한국데이터베이스진흥센터에서 실시하는 교육과정 안내입니다

- **교육명** : 인터넷 비즈니스 기획·설계·예산 관리 실무 과정
- **과정설명**
 - 인터넷 비즈니스의 전략 기획을 비롯해 예산 등에 걸친 문제 해결 능력 향상 과정
 - 인터넷 비즈니스 모델 기획 및 아이템 선정 훈련
- **교육일정 및 기간**
 - 일정 : 5월 14일(월) ~ 5월 18일(금)
 - 시간 : 13:00 ~ 17:00
 - 기간 : 5일(1일 4시간, 총 20시간)
- **교육비** : 29만원/1인

일차	교과명	교육내용
1일	인터넷 비즈니스 패러다임의 이해	· 인터넷 비즈니스의 뉴패러다임 · 인터넷 비즈니스의 모델 · 인터넷 비즈니스 모델별 사업 전략 · 인터넷 마케팅 전략 소개
2일	인터넷 비즈니스 시장조사 및 사업계획서 작성	· 시장조사 및 아이템 선정 · 경쟁업체 분석 · 내부역량 평가 · 목표시장 평가 · 일정 및 예산 · 사업계획서 작성
3일	웹사이트 제작 프로세서 계획	· 기획 및 전략 · 디자인과 스펙의 프로토타입 제작 · 제작 · 테스트 · 웹 사이트 제작 프로세스 관리
4일	인터넷 비즈니스 설계 기획	· 내용 설계 - 유형분석 - 웹 페이지 설계 - 설계 절차 및 지침 · 구조 설계 - 사이트 및 페이지 구조 - 설계 절차 및 지침 · 항해설계 - 핵심 항해 요인 - 기본 링크 유형 - 설계 절차 및 지침 · 화면 설계 - 기본 요소 - 설계 절차 및 지침
5일	인터넷 비즈니스 사후처리	· 최신 인터넷 마케팅 방안 및 사례 · 자금관리 방법

- **문의 및 신청**
 - 담당 : (재)한국데이터베이스진흥센터 교육센터
 - 주소 : 서울시 중구 다동 10번지 한국관광공사빌딩 9층
(영풍문고 맞은편, 1호선 종각 5번출구, 2호선 을지로입구 2번 출구)
 - 온라인 신청 : <http://www.dpc.or.kr>
 - 전화 : 3708-5371, 3708-5370
 - FAX 신청 : 318-5040(교육참가신청서 이용)